

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY,

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPCO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, OPENAI
HOLDINGS, LLC,

Defendants.

Case No. 1:23-cv-11195 (SHS) (OTW)

**REPLY MEMORANDUM OF LAW
IN FURTHER SUPPORT OF
DEFENDANT OPENAI OPCO,
LLC'S OBJECTION TO
NONDISPOSITIVE ORDER
PURSUANT TO FEDERAL RULE
OF CIVIL PROCEDURE 72(a)**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. ARGUMENT	1
A. The Order is Contrary to Binding Fair Use Precedent	2
1. The Order Improperly Limited Discovery Into Plaintiff's Conduct	2
2. The Order Contravenes Binding Precedent on the Market Harm Inquiry	5
3. The Order Precludes Highly Relevant First and Fourth Factor Discovery	8
B. The Order Ignores the Relevance to Damages and Other Remedies	9
III. CONCLUSION	10

I. INTRODUCTION

Plaintiffs in fair use cases are not immune from discovery into their own conduct. The Order challenged here effectively presumed the opposite. That result was contrary to law. The Times’s response is principally an exercise in misdirection: rather than defend the fundamental flaws in the underlying decision, the Opposition addresses arguments OpenAI did not make regarding discovery OpenAI is not seeking.

The Court should decline the Times’s invitation to ignore what is actually at issue—the underlying Order’s legal errors. This is not a garden-variety discovery dispute over what the Times has or has not actually produced. Rather, OpenAI is objecting to an Order (ECF 344) that severely cabins the scope of fair use discovery in a manner contrary to binding precedent. This Court’s intervention is necessary to correct the Order’s clear errors and ensure that this case—which presents questions of first impression involving new, revolutionary technology—is litigated on a complete and balanced record.

II. ARGUMENT

In this case, the Times takes aim at OpenAI’s creation of technology—generative AI large language models—that even the Times describes as having “transformative force” that could affect change “much like other technological revolutions.” *See Steve Lohr, How A.I. Could Reshape the Economic Geography of America, NY Times (Dec. 26, 2024), www.nytimes.com/2024/12/26/technology/ai-economy-workers.html.* Fair use is one of OpenAI’s primary defenses in this case. *See ECF 363 (“Obj.”) at 1.* The Times, for its part, contends that OpenAI’s alleged copying of the Times’s works in the creation of large language models constituted copyright infringement rather than fair use. *See id. at 2.* The Court will thus have to engage in the “multifaceted assessment” fair use requires, guided by the *nonexhaustive* list of factors in Section 107 of the Copyright Act. *Authors Guild v. Google, Inc.*, 804 F.3d 202, 213 (2d Cir. 2015) (“Google Books”);

see Google LLC v. Oracle Am., Inc., 593 U.S. 1, 18–19 (2021).

The Times disputes none of that. The Times nevertheless asks this Court to uphold an Order that denied OpenAI discovery critical to the fair use analysis that is needed to create a full record on the issue. *See Obj.* at 5–7; *see also* ECF No. 344 at 5–6. In so doing, the Times—like the Order below—ignores not only binding Supreme Court and Second Circuit case law on the scope of the fair use inquiry, *see Obj.* at 10–17, but also that “relevance, for purposes of discovery, is an extremely broad concept.” *Chen-Oster v. Goldman, Sachs & Co.*, 293 F.R.D. 557, 561 (S.D.N.Y. 2013) (citation omitted). This Court should reject the Times’s attempt to distract the Court from what is actually at issue here—a legally erroneous decision that threatens the entire trajectory of this case and thus requires this Court’s correction.

A. The Order is Contrary to Binding Fair Use Precedent

1. The Order Improperly Limited Discovery Into Plaintiff’s Conduct

As OpenAI explained in its Objection, the Order at issue held—contrary to law—that the statutory factors in Section 107, although “non-exclusive,” do *not* “require[] scrutiny” of—and thus do not permit discovery into—a copyright plaintiff’s own conduct related to the matters at issue. Order at 2–3. On the basis of that ruling, the Court refused to grant OpenAI critical discovery into “(1) the Times’s use of nonparties’ generative artificial intelligence (‘Gen AI’) tools; (2) the Times’s creation and use of its own Gen AI products; and (3) the Times’s position regarding Gen AI (e.g., positions expressed outside of litigation, knowledge about the training of third-party Gen AI tools using the Time’s works).” Order at 1, 5–6. In other words, the Order precludes OpenAI from obtaining discovery into whether and to what extent the Times’s own activities parallel the very actions it challenges here—the training and use of large language models.

The Times’s only response to this error addresses an argument OpenAI did not make.

According to the Times, the Court should reject OpenAI’s position because “OpenAI does not offer a single controlling case in which a court went beyond the statutory factors when resolving a fair use dispute.” ECF 429 (“Opp.”) at 7. The Times’s response reinforces why the Order below is incorrect. OpenAI did not have to identify cases in which a court “went beyond the statutory factors” to demonstrate the error here because the discovery at issue *is* relevant to the statutory factors themselves, as OpenAI explained. *See Obj.* at 11–13. Because “[e]ach factor . . . stands as part of a multifaceted assessment”—the assessment of whether a use qualifies as fair—the range of discovery relevant to the statutory factors *includes* discovery addressing a plaintiff’s actions and admissions related to the matters at issue in a case. *Google Books*, 804 F.3d at 213. That is why, in both *Oracle* and *Chegg*, the courts viewed as relevant what the plaintiff copyright owner had done and the positions the plaintiff copyright owner had taken. *See Oracle*, 593 U.S. at 31–32 (considering evidence that “[plaintiff] itself” had copied code in similar way to defendant, and what “[plaintiff] executives thought” about others’ uses); *Pearson Educ., Inc. v. Chegg, Inc.*, No. 21-cv-16866, 2023 WL 3775317, at *2, *6, *8 (D.N.J. June 2, 2023) (granting discovery request in copyright case for documents showing “industry custom and practice,” including plaintiff’s own use and plaintiff’s position on defendant’s copying).

Even taken on its own terms, the Times’s argument—like the Order below—is also incorrect in asserting that the fair use test is strictly limited to the enumerated statutory factors. As the Supreme Court made abundantly clear in *Google v. Oracle*, the fair use doctrine is “flexible,” 593 U.S. at 20, and Section 107’s “list of factors is *not* exhaustive,” *id.* at 19. Indeed, the fair use inquiry requires consideration of “relevant circumstances,” which can include—but are not limited to—matters that fall squarely within the statutorily enumerated factors. *See id.* at 19–20. Here, notwithstanding fair use’s flexibility and the multifaceted inquiry that is required,

the Order concluded that a copyright plaintiff's actions and positions related to the matters at issue in the case—if not directly tied to the defendants—are categorically irrelevant. Order at 2–4. That is not the law. *See, e.g., Oracle*, 593 U.S. at 31–32, 36.¹

Later in its response, the Times attempts to avoid the import of the Supreme Court's *Oracle* decision by again responding to an argument OpenAI did not make. According to the Times, *Oracle* does not support OpenAI's position because it did not “endorse[] the broad view that a plaintiff's activities are *always* relevant to the fair use inquiry.” Opp. at 12 (emphasis added). OpenAI never said otherwise, nor did it need to. What *Oracle* shows is that, contrary to the Order below, a plaintiff's statements and actions *related to the matters at issue* are relevant to the fair use inquiry. *See* 593 U.S. at 36 (considering Oracle's predecessor's “efforts to move into the [relevant] market”). The Second Circuit demonstrated the very same thing in *American Geophysical Union v. Texaco, Inc.*, where the court considered industry practices when evaluating the fourth fair use factor. 60 F.3d 913, 930 (2d Cir. 1994). The Order's conclusion here that, notwithstanding *Oracle* and *Texaco*, the fair use inquiry “do[es] not require court to examine” a plaintiff's actions or statements related to the matters at issue absent a direct connection to a defendant—and that discovery into such matters is thus categorically prohibited—was thus contrary to both *Oracle* and *Texaco*.

The Times's final attempt to evade the law of *Oracle* is equally futile. In the Times's view, because the Supreme Court in *Oracle* “was not ruling on whether the evidence was discoverable or admissible,” the Court's reliance on evidence of the sort OpenAI seeks here was “merely dicta.” Opp. 13. That is nonsensical. The Times does not dispute that the Supreme Court expressly relied on evidence of the *plaintiff's* own words and actions in conducting its fair use analysis, stating that

¹ Of course, the importance of a “flexible” fair use doctrine (*Oracle*, 593 U.S. at 20) and complete record is particularly important in cases of first impression like this one involving new, revolutionary technology.

“[t]hese and related facts *convince us* that the ‘purpose and character’ of Google’s copying was transformative—to the point where this factor too weighs in favor of fair use.” 593 U.S. at 31–32 (emphasis added). If the Supreme Court relied on this evidence in its holding, the same evidence cannot be irrelevant *as a matter of law* to a fair use defense.² The Times, in other words, simply asks this Court to ignore the Supreme Court’s analysis in *Oracle* and shield from discovery the “convinc[ing]” evidence OpenAI now seeks.

2. The Order Contravenes Binding Precedent on the Market Harm Inquiry

Although the Times spills much ink responding to OpenAI’s request for market-related discovery, the Times says little to actually defend the Order’s ruling, presumably because the Times recognizes that it is indefensible.

The Times starts by asserting that OpenAI forfeited its argument that the discovery sought is relevant to whether a licensing market exists. Opp. at 8–9. The Times is wrong. OpenAI argued in its pre-motion letter to Magistrate Judge Wang that the discovery sought is relevant to, among other things, whether there is a “workable market” for licensing data to create a large language model. ECF No. 236 at 3. And the Order rejected that argument (among others). See Order at 3 n.2, 5–6. There is no forfeiture here.

The Times’s arguments on the merits are equally unavailing. The Times claims that it “has not developed its own LLMs,” Opp. at 11, and that its use of its own or third-party generative AI tools is irrelevant to the impact of OpenAI’s alleged use of the Times’s works to train large language models “on the market for the Times’s works,” *id.* at 10. This ignores the threshold question under the fourth factor—whether there is a “ready market or means to pay for the use” at

² And even if the Supreme Court’s statement was somehow dicta, it “must be given considerable weight and can not be ignored,” *United States v. Bell*, 524 F.2d 202, 206 (2d Cir. 1975), as the Order did here. *See also Fuld v. Palestine Liberation Org.*, 82 F.4th 74, 100 (2d Cir. 2023) (noting that Supreme Court dicta is “accord[ed] deference . . . where . . . no change has occurred in the legal landscape”).

issue. *Texaco*, 60 F.3d at 931; see also *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614 (2d Cir. 2006). Here, that requires consideration of whether there is a “ready market” for data to train large language models.

The Times, like the Order below, simply ignores this threshold question and the relevance of the Times’s own actions and analyses to it. The Times chose to sue OpenAI for copyright infringement and to allege that developers of large language models “should first approach [rightsholders] about a licensing agreement.” FAC ¶ 49. Yet at the same time, the Times is deploying large language models in its business operations—models that the Times does not dispute were almost certainly trained on unlicensed data.³ Indeed, the Times does not even try to defend the Order’s clearly erroneous assumption that the other large language models it uses “presumably were developed without copying.” Order at 4. The Times has also indicated that it has worked to develop generative AI tools. See Obj. at 2–3.⁴ Having alleged that developers of such models should seek licenses, one would expect the Times to have evaluated whether it could itself develop or use large language models trained only on licensed data. This is important evidence for the fair use inquiry, and OpenAI is entitled to discovery into the Times’s evaluation of and conclusions about the impossibility of doing so. See Obj. 3–4, 14. The Order’s conclusion otherwise was error.

Indeed, the Times nowhere addresses that the Order’s conclusion is directly contrary to the Second Circuit’s decision in *Texaco*. As OpenAI explained, the Second Circuit in *Texaco*

³ The Times disingenuously claims that this fight is about documents related to “autocorrect and spell check.” Opp. 1. But OpenAI made very clear at the hearing before the magistrate that it was *not* seeking those types of documents, as the court acknowledged. See Sept. 12, 2024 Hr’g Tr. 112:21–25 (“So you are excluding all of that.”).

⁴ The fact that the Times claims not to be training its own large language models now, see Dkt. 429-2, does not obviate the need for the discovery at issue. In fact, it heightens it—if the Times, in the past, sought to develop such a model, but stopped after, for example, determining that there is no viable licensing market for training data, that would be highly relevant here. This is important evidence for the fair use inquiry, and OpenAI is entitled to discovery into the Times’s efforts to create such comparable tools and its reasons for abandoning any such efforts.

explicitly relied on evidence from parties other than the defendant to evaluate whether a viable market existed. *See Texaco*, 60 F.3d at 930 (2d Cir. 1994) (noting that “[t]he District Court found that many major corporations now subscribe” to “systems for photocopying licenses,” which suggests that “there currently exists a viable market for licensing these rights”). Under *Texaco*, the Times’s conduct and positions are unquestionably relevant and thus discoverable.

The Times likewise does not dispute that, contrary to the Order below, the Supreme Court in *Oracle* found relevant to its fourth factor analysis evidence of the plaintiff’s conduct unrelated to the defendant. *See Opp.* at 12–13; *see also Oracle*, 593 U.S. at 36. The Order, in contrast, held that such evidence is not discoverable because the fair use factors focus on the “*defendant’s* purported use.” Order at 2–3 (emphasis in original), 5–6. The Times tries to side step the irreconcilability of these decisions by focusing on the fact that the *Oracle* court was considering the second question in the fourth factor inquiry—whether harm to a viable market occurred. *See Opp.* at 13. But that misses the point. What matters is that the Supreme Court in *Oracle* found relevant to the fair use analysis the plaintiff’s own actions whereas the Order here held that discovery into such actions is categorically prohibited.

Rather than address head on what the Order actually holds, the Times warns of a parade of horribles that it says will result from allowing this discovery, including “extensive discovery” about “third-party misconduct.” Opp. 11–12. That is specious. Discovery into the Times’s own development or deployment of non-party generative AI does not require any determination regarding what third parties have or have not done. That is because OpenAI is focused on what the *Times* has said and done—(1) whether the *Times* evaluated the possibility of developing or using large language models trained only on licensed content, and (2) whether the *Times* concluded that doing so was impossible. This evidence is directly relevant to whether there is a “ready market

or means” through which OpenAI could have licensed the content required to train a large language model. *Texaco*, 60 F.3d at 929–31. And try as it might, the Times says nothing to establish otherwise.

3. The Order Precludes Highly Relevant First and Fourth Factor Discovery

The Times concedes that the transformativeness of the use at issue—here, the development of revolutionary large language models—is relevant to fair use, both the first factor, and the fourth factor (as to public benefits reaped from the use). Opp. at 14; *see also* Obj. 16–17. However, the Times argues that what is relevant to the transformativeness and public benefit inquiries is what *OpenAI* has done, not what the Times has done in connection with “nonparties’ generative AI tools.” Opp. at 15. As a result, the Times contends, the Order was correct to prohibit OpenAI from obtaining evidence from the Times about the Times’s own use of non-OpenAI large language models. *Id.* at 14–15.

Here, too, the Times simply ignores—as did the Order—the relevance of the Times’s own use of the same technology OpenAI developed, which the Times uses to improve and further its journalistic enterprise. Indeed, at the very same time the Times seeks to indict OpenAI’s creation of large language models, claiming that “GenAI Products Threaten High-Quality Journalism,” ECF 170 at 14; *see also id.* ¶¶ 47–54, the Times is publicly touting the transformative, revolutionary nature of the exact same technology for the Times’s own journalistic efforts. *See, e.g., Principles for Using Generative A.I. in The Times’s Newsroom*, NY Times (May 9, 2024), <https://www.nytco.com/press/principles-for-using-generative-a%E2%80%A4i%E2%80%A4-in-the-timess-newsroom/> (“Generative A.I. can assist our journalists in uncovering the truth and helping more people understand the world. Machine learning already helps us report stories we couldn’t otherwise, and generative A.I. has the potential to bolster our journalistic capabilities even more.”). Thus, discovery showing the extent and details of the Times’s use of the very

technology—large language models—that the Times has taken aim at in this case is directly relevant to the fourth fair use factor. *See Google Books*, 804 F.3d at 217 (“*Google Books*”) (tool that allowed readers to engage in new kinds of research was transformative, favoring fair use); *Oracle*, 593 U.S. at 39 (considering whether fair use holding would encourage “creative improvements, new applications, and new uses” in fourth factor analysis); *see also Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1523 (9th Cir. 1992) (use that “led to an increase in the number of independently designed video game programs” was fair). The Order’s refusal to grant such discovery disregarded that unquestionable relevance.

More troubling is the Times’s failure to tell the Court that the Times itself views as relevant the use of nonparty generative AI tools and the impact of the use of such tools on journalism. Less than two weeks before filing its opposition, the Times served a subpoena on Google seeking, among other things, “Documents concerning the behavior of users of [Google’s and Defendants’] Generative AI Products and Services” and “Documents concerning the impact of [Google’s and Defendants’] Generative AI Models and Generative AI Products and Services on Journalism.” *See* Ex. A (subpoena to Google). OpenAI is seeking the exact same categories of discovery from the Times—documents concerning The Times’s use of generative AI products and services, including Google’s, and documents concerning the impact of the use of Generative AI products and services, including Google’s, on journalism. *See* ECF 236-1 at 15 (RFP Nos. 45–49). Yet the Order refused to provide OpenAI exactly that, concluding that generative AI discovery unrelated to OpenAI is categorically off limits. *See* Order at 5–6. That conclusion is contrary to law. *See Oracle*, 593 U.S. at 31–32, 36; *see also Google Books*, 804 F.3d at 217.

B. The Order Ignores the Relevance of the Discovery Sought to Remedies

The Times does not dispute that the discovery OpenAI seeks regarding internal Times analyses is relevant to the objective reasonableness of OpenAI’s conduct, Opp. 15–16, which bears

on the appropriateness of various remedies the Times has sought. *See Obj.* at 20–21; *see also Chegg*, 2023 WL 3775317, at *6–8. Instead, the Times claims that it is providing such discovery. What the Times says it is providing, however, is only those analyses the Times has unilaterally decided are “relevant” and related specifically to OpenAI. The Times simply ignores that what is critical to the fair use inquiry, and what OpenAI entitled to—and what the Order wrongly denied—is discovery into any analyses by the Times of whether the training of large language models on unlicensed data is itself a fair use (regardless of whether the Times specifically referenced OpenAI in any such analysis). The Order’s conclusion that the Times may withhold such discovery reflects an unduly narrow interpretation of Federal Rule of Civil Procedure 26, which is contrary to law. *See Bryce Corp. v. XL Ins. Am., Inc.*, No. 1:23-cv-1814, 2024 WL 3219839, at *2 (S.D.N.Y. June 27, 2024) (quoting *Oppenheimer Fund, Inc. v. Sanders*, 437 U.S. 340, 351 (1978)).

III. CONCLUSION

For the foregoing reasons, OpenAI respectfully requests that this Court vacate the Order and order the Times to produce the requested discovery.

By: /s/ Elana Nightingale Dawson

LATHAM & WATKINS LLP

Andrew M. Gass (*pro hac vice*)

andrew.gass@lw.com

Joseph R. Wetzel

joseph.wetzel@lw.com

505 Montgomery Street, Suite 2000

San Francisco, CA 94111

Telephone: 415.391.0600

Sarang V. Damle

sy.damle@lw.com

Elana Nightingale Dawson (*pro hac vice*)

elana.nightingaledawson@lw.com

Michael David

michael.david@lw.com

555 Eleventh Street, NW, Suite 1000

Washington, D.C. 20004

Telephone: 202.637.2200

Allison L. Stillman

alli.stillman@lw.com

Rachel Renee Blitzer

rachel.blitzer@lw.com

Herman Yue

herman.yue@lw.com

Luke A. Budiardjo

luke.budiardjo@lw.com

1271 Avenue of the Americas

New York, NY 10020

Telephone: 212.906.1200

By: /s/ Joseph C. Gratz

MORRISON & FOERSTER LLP

Joseph C. Gratz (*pro hac vice*)*

jgratz@mofo.com

Vera Ranieri (*pro hac vice*)

vranieri@mofo.com

Rose S. Lee (*pro hac vice*)

rlee@mofo.com

425 Market Street

San Francisco, CA 94105

Telephone: 415.258.7522

Eric K. Nikolaides

enikolaides@mofo.com

Emily C. Wood
ewood@mofo.com
250 West 55th Street
New York, NY 10019-9601
Telephone: 212.468.8000

Carolyn Homer (*pro hac vice*)
chomer@mofo.com
2100 L Street, NW Suite 900
Washington, D.C. 20038
Telephone: 202.887.1500

By: /s/ Andrew F. Dawson

KEKER, VAN NEST & PETERS LLP

Robert A. Van Nest (*pro hac vice*)
rvannest@keker.com
Paven Malhotra
pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
mybarra@keker.com
Nicholas S. Goldberg (*pro hac vice*)
ngoldberg@keker.com
Thomas E. Gorman (*pro hac vice*)
tgorman@keker.com
Katie Lynn Joyce (*pro hac vice*)
kjoyce@keker.com
Christopher S. Sun (*pro hac vice*)
csun@keker.com
R. James Slaughter (*pro hac vice*)
rslaughter@keker.com
Edward A. Bayley (*pro hac vice*)
ebayley@keker.com
Andrew F. Dawson (*pro hac vice*)*
adawson@keker.com
Andrew S. Bruns (*pro hac vice*)
abruns@keker.com
633 Battery Street
San Francisco, CA 94111-1809
Telephone: 415.391.5400

Attorneys for OpenAI

*All parties whose electronic signatures are included herein have consented to the filing of this document in accordance with Rule 8.5(b) of the Court's ECF Rules and Instructions.

CERTIFICATE OF WORD COUNT COMPLIANCE

I, Elana Nightingale Dawson, an attorney admitted *pro hac vice* to practice before this Court, hereby certify pursuant to Local Civil Rule 7.1(c), and Rule 2.C of Judge Sidney H. Stein's Individual Practices, that the foregoing Reply Memorandum of Law was prepared in Microsoft Word and contains 3,391 words. In making this calculation, I have relied on the word and page counts of the word-processing program used to prepare the document.

Dated: January 27, 2025
New York, New York

/s/ Elana Nightingale Dawson
Elana Nightingale Dawson